Milk mid-infrared spectra based biomarkers contributing to genetic improvement for udder health, fertility and longevity

N. Gengler[1], A. Mineur[1], S. Vanderick[1], H. Hammami[1] & GplusE Consortium[2]

[1] ULiège-GxABT, Passage des Déportés 2, 5030 Gembloux, Belgium
[2] http://www.gpluse.eu (mark.crowe@ucd.ie)

Summary

Recent research showed the usefulness of using estimated breeding values (EBV) for mid-infrared (MIR) based biomarkers in genetic improvement. Similarly, research has also shown that genetic variation is contained in the absorbance traits along the MIR band of wavelengths. Targeted extraction of the useful genetic variance can be achieved by the combination of EBV. Direct estimation of EBV for absorbance traits was demonstrated. Our first objective was to show that the reduction of the rank of the (co)variance structure among spectral traits is possible by imposing linear functions, even if these functions represent lower accuracy MIR biomarkers. MIR based biomarkers traits were derived from ongoing research in the FP7 GplusE project. In this study, the pathway from MIR spectra to the use in genetic improvement will be described. First, blood reference phenotypic data was collected on Holstein cows, at early lactation for IGF-1, glucose, urea, cholesterol, fructosamine, β-hydroxybutyric (BHB) acid and non-esterified fatty acids (NEFA). These traits were calibrated against corresponding MIR spectral data. Calibration R^2_{cv} ranged from 0.21 to 0.51, very low from a chemometrical point of view, but potentially sufficient to extract useful spectral variation. This was validated, using EBV that were based on these MIR predictions for 144,623 records (closest to days in milk 25), from 73,378 cows, in the Walloon region of Belgium. Single-trait, but multi-lactation (1, 2, 3+) models yielded h2 estimates ranging from 0.07 to 0.27. At least 20 daughters with novel traits and official EBV for udder health, fertility and longevity with minimum reliabilities of 70% were required; a total of 124 bulls met this criteria. Standard selection index theory would usually rely on prediction error variance minimisation and estimated population (co)variances. Alternatively in this study, Partial Least Squares were applied to EBV for the milk MIR based biomarkers to develop novel genetic predictors, for udder health, fertility and longevity, by extracting genetic variation along the wave band after rank reduction. Using all bulls, correlations between best predictors and EBV for udder health, fertility and longevity were at least 0.63, 0.67 and 0.62. Using selection index theory and based on significant increases of prediction abilities of longevity (0.76 compared to 0.68 from udder health or fertility alone) using also milk MIR based blood biomarkers, their potential contribution to genetic improvement of udder health, fertility and longevity will be demonstrated.

Keywords: milk MIR spectra, blood biomarkers, genetic improvement, udder health, fertility, longevity

Introduction

Genetic improvement for functional and health traits depend on the availability of relevant data (Egger-Danner et al., 2014). Recent efforts allow the development of appropriate genetic evaluation systems for udder health, fertility and longevity in many breeds and population. However given the specific context of these traits achieving earlier high reliabilities can be problematic. For example, fertility can only be assessed later during lactation, udder health relies often on somatic cells, or on unreliable mastitis records and longevity of a cows is truly known at her death. For this reason, alternative, early indicator traits could be considered useful under the hypothesis that they are available very early during the productive life of a given cow.

Milk mid-infrared (MIR) spectral data has been identified as a major potential source of relevant data through its easy, cheap and routine use in milk analysis. Calibration of milk based biomarkers has been reported by several authors. In a more limited fashion MIR predicted milk based biomarkers start also to be used. However reference traits are blood biomarkers. Therefore the European FP7 Project GplusE (http://www.gpluse.eu) has collected blood biomarkers for insulin growth factor 1 (IGF-1), glucose, urea, cholesterol, fructosamine, β-hydroxybutyrate (BHB) and non-esterified fatty acids (NEFA). Efforts to calibrate those biomarkers against milk MIR spectral data are ongoing (C. Grelet, personal communication) but MIR predicted blood biomarkers are not expected to be reliably predicted such. In the context of this study used predictors were based on R^2_{cv} of 0.21 to 0.51, very low from a chemometrical point of view, but potentially sufficient for the extraction of useful spectral variation.

The objective of this study was to test if the targeted combination of estimated breeding values (EBV) for biomarkers predicted with low accuracies from MIR spectra, increased their usefulness in the genetic evaluation of dairy cows for udder health (UDH), fertility (FERT) and longevity (LONG).

Material and methods

All research was based on data collected in the Walloon region of Belgium. Spectral MIR records were collected since 2012 and standardised. Using the available equation IGF-1, glucose, urea, cholesterol, fructosamine, BHB and NEFA were predicted from MIR data. for 144,623 records (test-day closest to days in milk 25), from 73,378 cows. Single-trait, but multi-lactation (1, 2, 3+) models were fitted and yielded h^2 estimates ranging from 0.07 to 0.27. Official EBV for UDH, FERT and LONG were obtained from the official genetic evaluation in the Walloon Region of Belgium (www.elinfo.be). Some of the EBV for these bulls were from MACE genetic evaluations by INTERBULL. At least 20 daughters for each bull with novel trait data were required. Moreover, novel traits and UDH, FERT and LONG had to have a minimum reliability (REL) of 70%. A total of 124 bulls met these criteria. By imposing rather strict REL, observed correlations between EBV for different traits were expected to be closer to genetic correlation avoiding the need to use the correction proposed by Calo *et al.* (1973). Standard selection index theory would usually rely on prediction error variance minimisation and estimated population (co)variances. However in this specific data high to extremely high correlations between the 21 biomarker EBV leading to major multicollinearity issues. Therefore, alternatively in this study, Partial Least Squares (PLS) were applied to EBV for the MIR based biomarkers to develop novel genetic predictors, for

udder health, fertility and longevity, by extracting genetic variation along the wave band after rank reduction. Cross-validation was done by choosing 10% randomly and predicting from the 90% other records, a process that was repeated 1000x. The Proc PLS procedure from SAS was used to do the computations.

Results and discussion

Reliabilities

Table 1. Heritabilities and average reliabilities (REL) in percent for the MIR predicted blood biomarkers reported for the 124 selected sires.

	Lactation 1		Lactation 2		Lactation 3+	
Blood biomarker	Average REL	h^2	Average REL	h^2	Average REL	h^2
IGF-1	92.0	0.21	93.0	0.26	92.7	0.24
Glucose	92.8	0.27	92.9	0.27	92.7	0.25
Urea	80.6	0.07	80.7	0.08	81.0	0.07
Cholesterol	85.1	0.09	82.3	0.08	88.9	0.17
Fructosamine	92.6	0.25	93.2	0.27	93.3	0.27
BHB	90.9	0.20	89.3	0.16	89.8	0.16
NEFA	88.0	0.14	88.6	0.13	88.9	0.12

As given in Table 1, based on the restrictive selection of bulls average REL were for all traits over 80%, the lowest values were reported for MIR predicted Blood Urea, a trait with low heritability (0.07 – 0.08). For the highest heritable traits the average REL were close to those found for LONG, FERT and UDH (93.2%, 93.8% and 97.8%) with h^2 of 0.11, 0.04 and 0.14.

Correlations

Table 2. Correlations for the MIR predicted blood biomarkers with udder health (UDH), fertility (FERT) and longevity (LONG) reported for the 124 selected sires.

	Lactation 1			Lactation 2			Lactation 3+		
Blood biomarker	UDH	FERT	LONG	UDH	FERT	LONG	UDH	FERT	LONG
IGF-1	0.05	0.18	0.14	0.07	0.18	0.16	0.18	0.25	0.22
Glucose	0.06	0.23	0.15	0.09	0.24	0.17	0.20	0.31	0.23
Urea	0.13	0.21	0.14	0.11	0.07	0.14	0.17	0.16	0.15
Cholesterol	-0.08	-0.06	0.09	0.02	0.03	0.20	0.05	0.03	0.20
Fructosamine	0.18	0.36	0.24	0.19	0.35	0.24	0.26	0.39	0.25
BHB	-0.01	-0.15	-0.09	-0.02	-0.12	-0.06	-0.08	-0.12	-0.07
NEFA	-0.12	-0.31	-0.23	-0.17	-0.30	-0.23	-0.24	-0.32	-0.25

Correlations between EBV for individual biomarkers in a given lactation ranged in absolute values between 0.01 (BHB and UDH in first lactation) and 0.39 between (Fructosamine and FERT in 3+ lactation (Table 2). Generally the values were rather low

compared to the theoretical links between blood biomarkers and UDH; FERT and in consequence also LONG. However one should not forget that we based these EBV on predictions with rather low R^2_{cv}.

Predicting udder health, fertility and longevity from milk MIR blood biomarkers

Table 3. Centred and scaled PLS (number of latent variables = 13) coefficients for the MIR predicted blood biomarkers with udder health (UDH), fertility (FERT) and longevity (LONG) reported for the 124 selected sires.

	Lactation 1			Lactation 2			Lactation 3+		
Blood biomarker	UDH	FERT	LONG	UDH	FERT	LONG	UDH	FERT	LONG
IGF-1	0.30	0.12	-0.12	-0.93	-1.35	-1.30	0.37	0.40	0.78
Glucose	-0.51	-0.26	-0.54	-0.02	0.15	0.17	1.42	1.68	1.74
Urea	0.15	0.39	0.00	0.01	-0.18	0.21	-0.03	0.01	-0.01
Cholesterol	-1.53	-1.14	-1.19	0.60	0.98	0.88	0.74	-0.04	0.33
Fructosamine	-0.01	0.18	0.24	-0.63	-0.44	0.85	0.11	-0.39	-1.68
BHB	0.66	0.27	0.28	-0.18	-0.53	-0.12	0.25	1.14	0.52
NEFA	-0.85	-0.56	-0.83	0.12	0.08	0.25	0.12	-0.34	-0.02

One feature of the use of PLS is its ability to deal with multicollinearity. When comparing regression coefficient obtained by ordinary least square regression (results not shown) and those reported here (Table 3), despite the very high genetic correlations between lactations (results not shown) for the same biomarkers, we see that only very few coefficients showed the expected strong opposition that was found in ordinary least square regression.

Predictions of UDH, FERT and LONG obtained using the coefficients in Table 3 explained 0.41, 0.45 and 0.39 of the total variance (R^2_{cv}). The associated cross-validation regression coefficients were 0.64, 0.67 and 0.62. In following paragraphs these predictions using milk MIR based blood biomarkers will be called pUDH, pFERT and pLONG (Figures 1, 2 and 3).

Different predictions of longevity

An important breeding goal trait is longevity also called herd life, or with a slightly different definition, survival. Unfortunately it is also a very difficult trait as relevant data is late available. Predictor traits are very commonly used and the result of this prediction called indirect longevity or herd-life (Jairath et al., 1998). Different types of predictors are commonly used, ranging from type traits to functional traits. In this study we have put into perspective the use of usual traits as functional traits UDH and FERT, but also the novel traits. Therefore pLONG was added as an alternative predictor.

Results based on the 124 sires, PLS and a cross-validation strategy to retain results with the optimum number of latent variables, are in Table 4. Results based on R^2_{cv} showed that the combination of two traits in the predictor was always a better choice than using UDH, FERT or pLONG alone. Moreover combining the three was based on this criteria the best choice with an approximate genetic correlation between direct and indirect longevity of 0.76. Most studies reported lower values.

However, high correlations are not directly sufficient to compare predictors. Another point is the information content (linked to data and heritability) that indirect longevity contains. In this study, average REL were used to represent this (Table 4.). Results showed once more that the addition of pLONG has a substantial effect.

Table 4. Results under a cross-validation scenario obtained for different predictors of longevity (indirect longevity) based on udder health (UDH), fertility (FERT) and predicted longevity (pLONG) obtained from milk MIR predicted blood biomarkers (results reported for the 124 selected sires).

Traits used in the predictor	r_{cv}	R^2_{cv}	LV[1]	REL_PRED[2]	REL_ORG[3]
UDH	0.60	0.36	1	35.4	97.8
FERT	0.68	0.46	1	43.2	93.8
pLONG	0.62	0.38	1	32.8	86.5
UDH, FERT	0.72	0.52	1	49.7	95.7
UDH, pLONG	0.70	0.49	1	44.9	92.0
FERT, pLONG	0.74	0.55	2	49.8	90.9
UDH, FERT, pLONG	0.76	0.57	1	53.2	92.7

[1] LV = latent variable
[2] REL_PRED = theoretical average reliability in % associated to this (combination of) trait(s) when used as predictor of longevity
[3] REL_ORG = theoretical average original reliability in % associated to this (combination of) trait(s)

Conclusions and perspectives

The innovative use of PLS instead of classical selection index type regressions proved to be beneficial, limiting the effect of multicollinearity. There are indeed still opportunities to improve the selection index theory, especially when predictor traits in the information vector are extremely highly correlated.

This research showed that even lower accuracy milk MIR based blood biomarkers can make a useful contribution in the estimation of EBV for important breeding goal traits. The targeted combination of EBV for these milk MIR predicted blood biomarkers increased their potential contribution to breeding goal traits, and therefore, their usefulness for genetic evaluation. This research demonstrated this in the context of longevity. Even if the gains observed through approximate reliabilities of indirect longevity remain rather modest, one should not forget that MIR data is available very early, as soon as the cow is milked. This study used MIR spectral data from the first test-day on, so potentially for a two year old. This should have a relevant impact on genetic improvement.

The definition of milk MIR predicted blood biomarkers may still be preliminary. Currently, improvements of their calibrations are foreseen (C. Grelet, personal communication). Furthermore, the genetic model used to analyse the 7 blood biomarkers could be changed to a full 21 traits model. Alternative phenotype definitions could also help to better extract information from spectra. Moreover, it might be suboptimal to first do a phenotypic calibration from MIR spectra, then genetic evaluations on these calibrated traits. A direct massive multi-trait genetic model taking absorbances directly for given MIR wavenumbers as traits (Soyeurt *et al.*, 2010) could be a better approach. However rank reduction techniques will be necessary to handle the very large number of traits.

Acknowledgements

The authors acknowledge the support of the whole GplusE team, of the Walloon Futurospectre Consortium (including also the Walloon Breeding Association, the Walloon Agricultural Research Centre and the Walloon Milk Committee) and of the European Milk Recording (EMR), providing access to MIR data standardization.

The GplusE project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration, under grant agreement n° 613689. The views expressed in this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

List of References

Calo, L. L., McDowell, R. E., Van Vleck, L. D., & Miller, P. D. (1973). Genetic aspects of beef production among Holstein-Friesians pedigree selected for milk production. J. Anim. Sci., 37(3), 676–682.

Egger-Danner, C., Cole, J. B., Pryce, J. E., Gengler, N., Heringstad, B., Bradley, A., & Stock, K. F. (2014). Invited review: Overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. Animal, 9(2), 191–207.

Jairath, L., Dekkers, J., Schaeffer, L. R., Liu, Z., Burnside, E. B., & Kolstad, B. (1998). Genetic Evaluation for Herd Life in Canada. Journal of Dairy Science, 81(2), 550–562.

Soyeurt, H., Misztal, I., & Gengler, N. (2010). Genetic variability of milk components based on mid-infrared spectral data. Journal of Dairy Science, 93(4), 1722–1728.

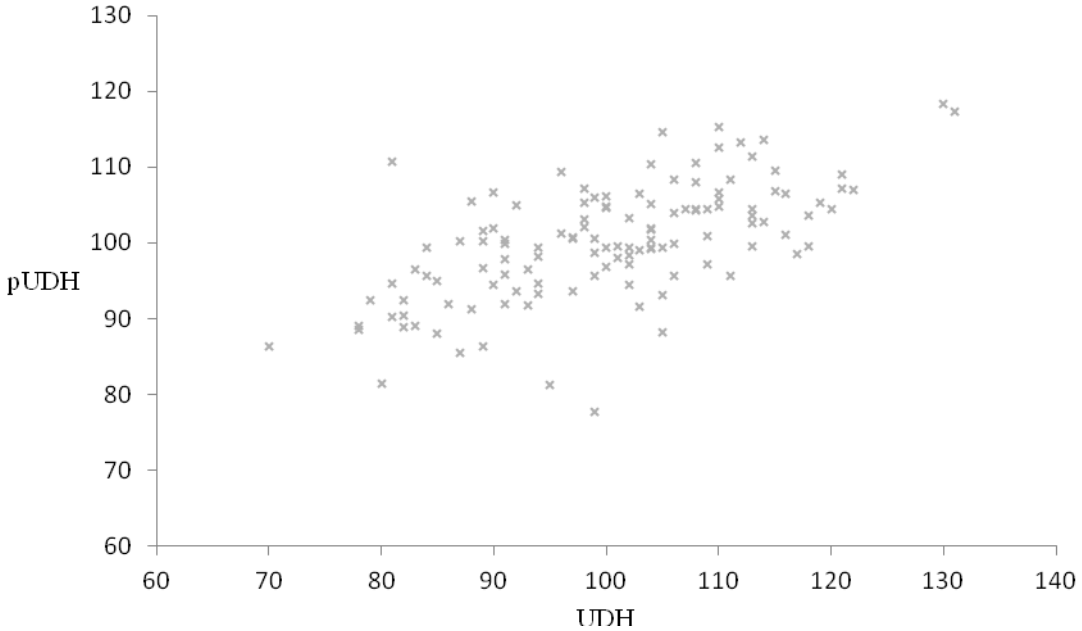

Figure 1. EBV for udder health (UDH) and predicted EBV udder health (pUDH) using milk MIR blood biomarkers (correlation between UDH and pUDH 0.63)

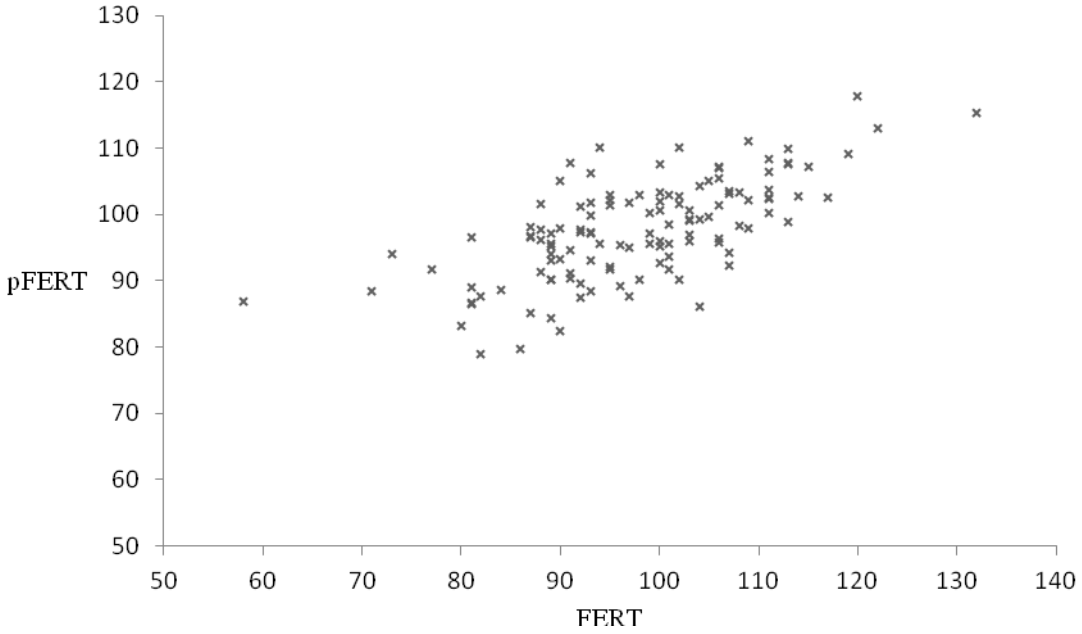

Figure 2. EBV for fertility (FERT) and predicted EBV fertility (pFERT) using milk MIR blood biomarkers (correlation between FERT and pFERT 0.67)

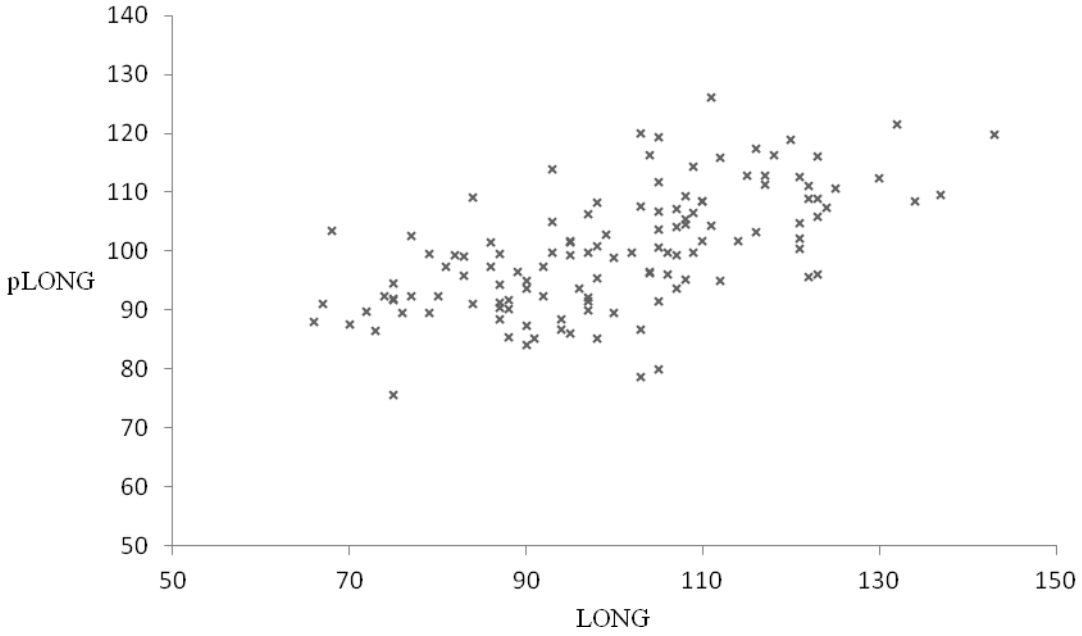

Figure 3. EBV for longevity (LONG) and predicted EBV longevity (pLONG) using milk MIR blood biomarkers (correlation between LONG and pLONG 0.62)